

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 18-03-2011		2. REPORT TYPE Journal Article		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Component Identification in Multi-Chemical Mixtures with Swept-Wavelength Resonant-Raman Spectroscopy				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Robert Lunsford, David Gillis, Jacob Grun, Pratima Kunapareddy, Charles Manka, Segei Nikitin				5d. PROJECT NUMBER BA09DET019	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Navy Research Laboratory, Washington, DC Research Support Instruments, Lanham, MD				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Threat Reduction Agency 8725 John J Kingman Road Fort Belvoir, VA 22060				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution Statement A: Approved for Public Release; distribution is unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The Swept Wavelength Optical Resonant Raman Detector (SWOrRD) at the Naval Research Laboratory is scanned through the 220 nm to 260 nm window of the ultraviolet and generates two dimensional spectral maps allowing for rapid identification of chemical materials. We create a library of chemicals whose Raman spectral features partially overlap and then measure the mixtures created by combinations of the library set. The library and mixtures are identically processed and loaded into a detection program. This system, utilizing a linear combination model enables the deconvolution of the multi chemical liquid mixture and a reconstitution of the fractional molecular abundances.					
15. SUBJECT TERMS Swept Wavelength Optical Resonant Raman Detector, SWOrRD, Ultraviolet, spectral maps					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 13	19a. NAME OF RESPONSIBLE PERSON Bryan Horner
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code) 703.767.3379

COMPONENT IDENTIFICATION IN MULTI-CHEMICAL MIXTURES WITH SWEPT-WAVELENGTH RESONANT-RAMAN SPECTROSCOPY

Robert Lunsford¹, David Gillis², Jacob Grun¹, Pratima Kunapareddy³, Charles Manka³, Sergei Nikitin³

¹ Plasma Physics Division, US Naval Research Laboratory, Washington DC,

² Remote Sensing Division, US Naval Research Laboratory, Washington DC,

³ Research Support Instruments, Lanham MD

Abstract

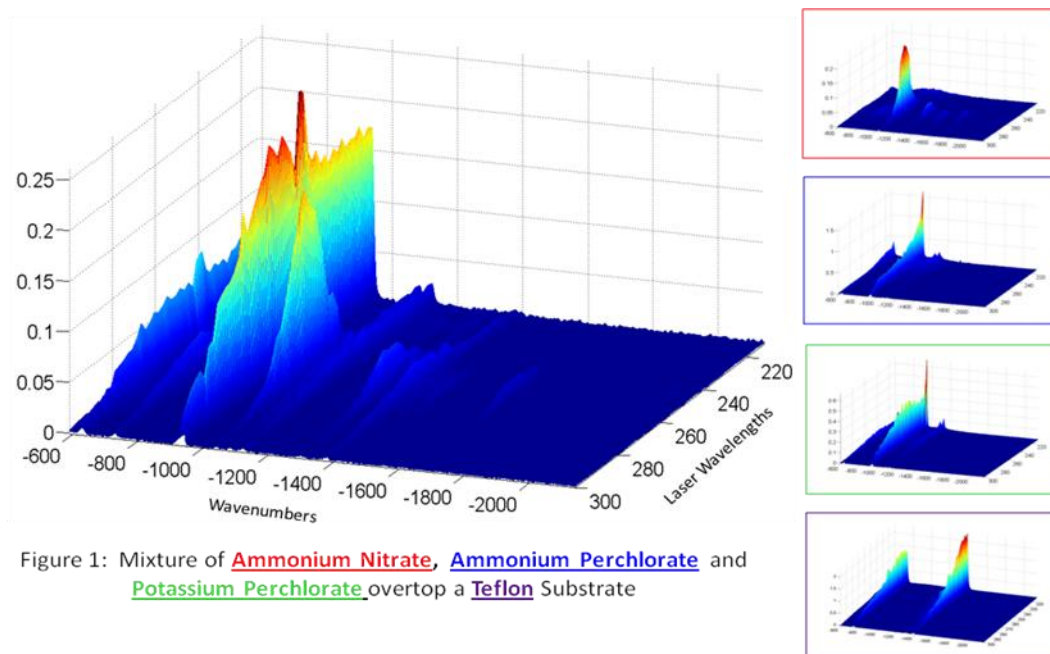
The Swept Wavelength Optical Resonant Raman Detector (SWOrRD) at the Naval Research Laboratory is scanned through the 220 nm to 260 nm window of the ultraviolet and generates two dimensional spectral maps allowing for rapid identification of chemical materials. We create a library of chemicals whose Raman spectral features partially overlap and then measure the mixtures created by combinations of the library set. The library and mixtures are identically processed and loaded into a detection program. This system, utilizing a linear combination model enables the deconvolution of the multi chemical liquid mixture and a reconstitution of the fractional molecular abundances.

Introduction

The utilization of Raman spectroscopy, specifically Ultraviolet Resonance Raman spectroscopy for the detection and identification of chemical, biological and nuclear hazards is of great interest due to the sensitivity and specificity afforded by this technique. Detection by means of optical probing is a fast, non-contact method which requires little to no sample preparation and can be performed remotely by an operator. In addition this method is also well suited for use by an automated monitor or a mobile autonomous system, such as an in-situ environmental detector or a sensor equipped unmanned vehicle. In addition, this technique offers the ability for a single detector to identify multiple species of targets. Many of the detectors presently available in this arena are extremely successful while operating within their narrow focus and yet are blind to alternate but equally deadly threats. However, for Raman

detection to become practicable as a forensic tool, the method must demonstrate the ability to distinguish the target elements while operating in a complex environment.

By employing a variable frequency laser source to illuminate the analyte we can interrogate the molecules at a number of distinct wavelengths. The compilation of these single wavelength illuminations generates a 2-Dimensional contour map of the functional form $I = f(x,y)$ where x and y are the excitation wavelength and the Stokes Raman frequency shift respectively. A typical spectral set of a mixture and its subsequent basis components is shown in Figure 1



In addition to the Raman peaks present in the single spectrum which provide information about the target compound's molecular structure, the two-dimensional signature also contains a record of the changes in the molecular resonance cross sections as the sample responds to variations in the illuminating laser wavelength. This additional dimensionality should create a signal which is much more robust and therefore harder to confuse, resulting in better specificity. This is particularly important when utilizing the device in realistic environments containing multiple target materials and contaminants.

The Swept Wavelength Optical Resonant Raman Detector (SWOrRD) at the Naval Research Laboratory is a stimulated spectroscopy system capable of rapidly acquiring the spectral signatures of both solid and liquid samples over a large range of laser wavelengths. The uniqueness of the SWOrRD system is its ability to both scan over a large range of wavelengths as well as to selectively tune to

wavelengths which overlap with the resonant energy bands of the target molecules. As it is impossible to completely determine a priori which wavelengths will be the most effective when attempting to detect a particular signal within a complex and rapidly changing environment, the operation of a system which is flexible in this regard holds distinct advantages over fixed wavelength systems.

In the work presented here, the 2-Dimensional Raman signatures of selected chemicals are measured and then utilized to form a detection basis. These chemicals are then combined in various permutations so as to create unknown mixtures which are then likewise measured. Upon completion of the full experimental run the raw spectra are Fourier filtered to remove both high-frequency noise and baseline instrument contributions and then recalibrated to compensate for power variation, sample absorption, wavelength drifts, system transmission and detector responses. They are then assembled to obtain a single multi-run spectrum which is treated as a single unit.

Utilizing a LabVIEW based detection program the components of the mixture are rapidly determined through an examination of the signal correlation between the unknown signature and those signatures stored in the chemical library. Once the mixture components have been identified an estimation of the relative abundances of the tagged chemicals is undertaken by comparison of the functional distance between the unknown signal and a weighted linear superposition of the library elements.

Experimental Apparatus

The primary components the SWOrRD system consist of a tunable laser and a coupled dual stage spectrometer with their associated diagnostics, computers and optics. The target is illuminated by a frequency doubled optical parametric oscillator pumped by the 3rd harmonic of a Nd-laser. The laser which operates at 1 kHz and generates 5 ns pulses is capable of the serial generation of wavelengths in three distinct windows; 210nm to 320 nm, 400 nm to 640 nm, and 710nm to 2100 nm in steps as small as 0.1nm. These windows are a result of the harmonic bands available to the OPO. The average laser power in each of these windows is dependent upon the selected wavelength and the conversion efficiency of the OPO, but ranges from up to 15 mW on target in the UV to 50 mW in the visible. This ability to illuminate a target with a broad range of laser wavelengths at a high average but low peak power allows for rapid interrogation and data collection with a minimum of collateral sample damage.

While active, the laser wavelength is monitored by a wavemeter which has itself been calibrated by a NIST traceable frequency stabilized laser source. The laser power is also measured prior

to each spectrum being obtained and the exposure times for the CCD camera are automatically adjusted to maintain constant sample illumination at each wavelength. There is also an inline laser power meter which records the laser output during the exposure so that minor fluctuations in energy on target can be corrected for in post processing.

The other major component of the SWOrRD system is the coupled data collection system. For the liquid samples utilized in this set of experiments light scattered from the illuminated sample is collected at an angle of 90 degrees with respect to the incident laser beam and the resultant scattered light is focused onto the entrance slit of an Acton double spectrometer consisting of two 0.5 meter stages with a selection of gratings and an attached Pixas CCD camera. The camera has been back illuminated and coated for enhanced UV response.

The run file which automates the collection process uses several input parameters to maintain a consistent experimental parameter space throughout the data collection. As noted previously, the transition between wavelengths within the laser system is accompanied by a variation in the laser output power due to nonlinearities within the efficiency of the wavelength conversion process. To compensate for this the power output of the laser is measured prior to each separate exposure. From this measurement an exposure time is calculated which approximately maintains the energy imparted to the sample across the illumination range. In addition to this pre-exposure measurement there is also an inline power meter which records the actual power output conveyed by the laser during the exposure. As the recording camera settings cannot be adjusted once the exposure trigger is activated and the shutter is opened, this integrated power value is stored along with the camera data in the requisite output file. Prior to importing the data into the processing toolbox, these variations in exposure are compensated for by numerically adjusting the recorded signal counts by the ratio between the requested exposure and the total power exposure recorded by the integrated power meter. This step is being integrated into the next version of the SWOrRD processing toolbox and should no longer represent a separate pre-processing step.

The spectrometer, camera, laser, beam steering mirrors, shutters for the laser and camera, as well as the power and wavelength meters are all connected to a control computer which is running a SWOrRD specific LabVIEW based control software. A signature is acquired by first retrieving an input file in which is explicated the desired experimental parameters. After tuning the laser to the requested wavelength, a power meter reading is taken and the desired exposure time is calculated. At this point the grating position is shifted so as to maintain an approximately constant location of the non-shifted

Rayleigh line within the spectrometer. Upon completion of a single wavelength exposure, the experimental parameters located on successive lines within the run file are loaded into the system and the subsequent runs proceed automatically.

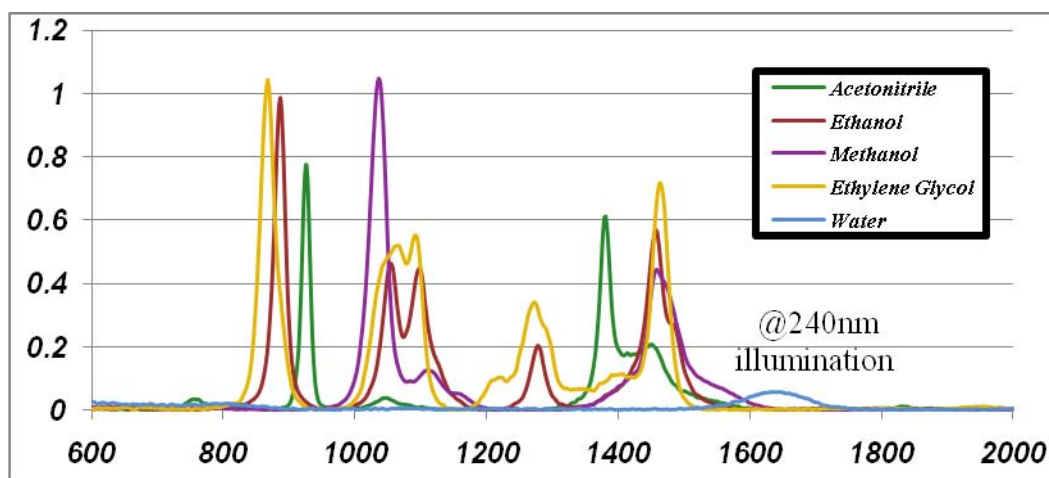
Materials

In these experiments we selected the five non-reactive high purity chemicals listed in table 1 to compromise the basis of our mixture set. These chemicals were chosen because their Raman spectral features partially overlap as is shown in Figure 2 so as to complicate the detection process. A sixth chemical (Cyclohexane) was also acquired and its signature recorded and added to the library. This chemical was not included as a component in any of the mixtures and its addition into the spectral library is solely for the purpose of determining null result accuracy.

Each mixture was created by combining equal volumes of the library chemicals into a constantly agitated graduated cylinder. A sample of the resultant mixture was then analyzed by a Agilent spectrophotometer to determine absorbance characteristics of the liquid. The remaining mixture was then placed into a standard UV transparent cuvette and illuminated by the laser system elucidated in the previous section.

SWOrRD Toolbox Processing

Upon completion of a multiwavelength experimental run the data and associated run files for that particular experiment are transferred from a collection drive which is attached to the system control computer onto a separate processing workstation. Once the data has been exposure compensated it is read into an NRL generated processing toolbox. This sequential step HDF based processing toolbox Fourier filters the raw spectra to remove both high-frequency noise and baseline instrument contributions and then recalibrates to compensate for power variation, sample absorption, wavelength drifts, system transmission and detector responses. At each step the data is saved in an HDF compatible format



Chemical Constituents: Need to add far wavelength spectra on separate scale to graph and confirm consistent color scheme

Chemical	Supplier	Molecules per mL	Select Raman Lines (cm ⁻¹)							
Ethanol	Warner-Graham 200 Proof	1.031E+22	886	1054	1097	1279	1457	2888	2936	2980
Methanol	Sigma-Aldrich HPLC Grade	1.488E+22	1036	1458	2840	2948				
Ethylene Glycol	Sigma-Aldrich Spectrophotometric Grade	1.080E+22	867	1272	1464	2885	2940			
Acetonitrile	Sigma-Aldrich HPLC Grade	1.153E+22	925	1380	2258	2298	2946	3006		
Water	Fluka HPCE Grade	3.344E+22	1640	3480						
Cyclohexane	Sigma-Aldrich HPLC Grade	5.574E+21	802	1028	1158	1266	1347	2664	2853	2925

Table 1: Selected Chemicals

enabling full traceability of the final data as well as back compatibility within the intermediate processing steps. These signatures are stored along with their relevant documentation and algorithm descriptions in a single library file. For this experiment the finished data was then converted back into a spreadsheet file structure allowing for greater data portability to other analysis programs.

Data manually aligned and buffered to create common signature basis

The Raman signal strength data is now in a full 2-dimensional array where the respective dimensions are input laser wavelength and wavenumber shift. As the motivation for the experiment is to determine the efficacy of the linear mixture model, the results of the automated alignment algorithm within the toolbox were manually verified to make certain that it had resulted in an optimized spectrum. Minor variations in the line shape and spectral dispersion which occur as a result of the variations in the laser wavelength can disrupt the alignment program and make this a necessary step.

To ensure consistency of the detector method these errors were corrected as follows. The 921 cm^{-1} and 2946 cm^{-1} Acetonitrile lines were chosen as alignment standards and these peaks were manually checked to ascertain their location across the collected spectra and were adjusted as necessary. Thus through all the mixtures which contained Acetonitrile, these two lines have been placed at these locations and the locations of the other lines are adjusted with respect to these values. For those mixtures which did not contain an Acetonitrile component, alternate isolated Raman lines were benchmarked within an Acetonitrile containing mixture and then these line locations were utilized as alignment standards for the aforementioned Acetonitrile lacking mixtures. The attempt here was to maintain a completely self consistent mixture set given our experimental apparatus without over determining the system with multiple reference lines from disparate sources.

Analysis and Results

Linear Mixture Model and Fractional Abundance Algorithm

Identification of components within the mixture was achieved by assuming a Linear Mixture Model. In this case, the signal of the resultant mixture is assumed to be a linear combination of the library signatures (Y_n). These signatures are individually weighted by an intensity α_n . Here the quantity α should be directly related to the volumetric mixture concentration by the molecular density and Raman cross section and our goal should be to minimize the fraction of signal relegated to the error term ω_{err} .

$$\text{Target Signal} = \alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_3 Y_3 + \alpha_4 Y_4 + \dots \alpha_n Y_n + \omega_{err}$$

To rapidly determine the presence or absence of a particular library element within an unknown mixture we perform an iterative maximization process. The Pearson correlation coefficient (r) of the sample is determined by comparing the covariance of the signal array from the unknown sample (\mathbf{X}) to each of those signal arrays within the library(\mathbf{Y}). Here N refers to the size of the array.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

The Library signature found to have the greatest correlation is determined to be the primary mixture component and weighted at unity. This signature is then tagged as a likely mixture component and the other library signatures are added to the primary component in strengths varying from zero to unity and again the correlation coefficient is maximized. Through this process a secondary mixture component is added and weighted with respect to the primary component. This pair of signals is then combined to form the new base signal to which additional library signatures are added. This process is iterated until the algorithm determines that no further maximization of signal correlation is possible through the addition of supplementary signatures. The results of this algorithm being run on our 5 chemical mixture set are shown in Chart 1. With this full and completely self consistent data set, the detection algorithm has performed extremely well. There were a few cases where the program detected a trace presence of a chemical which was not in the mixture and one case where the detection algorithm missed a mixture component.

Once the components of the mixture were determined, we then attempt to recreate the fractional molecular abundances of the constituent chemicals. While the adjusted correlation coefficient method performed well in the identification of the constituent chemicals and was rapid enough for one to foresee its implementation in a fieldable diagnostic, it was not successful in accurately determining the molecular concentrations of the sample. To recreate the mixture itself we rely on the most literal definition of a linear superposition and add together the library components to attempt a recreation of the unknown signal. Thus each of the α_i can take on a scaled weight between 0 (absence of component) and 100 (full signature presence) with the further constraint

Equal Volume Mixtures of		Acetonitrile	Methanol	Ethanol	Ethylene Glycol	Water	Cyclohexane
Acetonitrile							
Methanol							
Ethanol							
EthyleneGlycol							
Water							
Cyclohexane							
Acetonitrile and Methanol							
Acetonitrile and Ethanol							
Acetonitrile and Ethylene Glycol							
Acetonitrile and Water							
Methanol and Ethanol							
Methanol and Ethylene Glycol							
Methanol and Water							
Ethanol and Ethylene Glycol							
Ethylene Glycol and Water							
Ethanol, Methanol, and Acetonitrile							
Ethylene Glycol, Methanol, and Acetonitrile							
Methanol, Acetonitrile, and Water							
Ethanol, Ethylene Glycol, and Acetonitrile							
Ethanol, Acetonitrile and Water							
Ethylene Glycol, Acetonitrile and Water							
Ethanol, Methanol and Ethylene Glycol							
Ethylene Glycol, Methanol, and Water							
Ethanol, Ethylene Glycol and Water							
Ethanol, Ethylene Glycol, Acetonitrile and Water							
Ethylene Glycol, Methanol, Acetonitrile, and Water							
Ethanol, Acetonitrile, Water, and Methanol							
Ethanol, Methanol, Acetonitrile, and Ethylene Glycol							
Ethanol, Acetonitrile, Water, Methanol and Ethylene Glycol							

CORRECT DETECTION	CORRECT NEGATIVE	INCORRECT TRACE DETECTION	INCORRECT NEGATIVE
----------------------	---------------------	------------------------------	-----------------------

Table: Identification Results using correlation coefficient scoring

that the sum of the α components is equal to 100. As the simple brute force methodology of attempting all possible combinations of component weights is not feasible even with as limited a library scope as we are dealing with we created an iterative algorithm which searches a subset of the multi-dimensional space. It checks along the component weight axes by +/- 2 steps in each direction for each library component and then after choosing the point with the greatest agreement, recenters the algorithm and restarts the procedure. The assumption we are operating under, which has been born out to a large extent, is that the least squares difference function $f(x)$

which we are using to measure the distance between the unknown signal and the numerical recreation is smooth and fairly well behaved thus allowing for the local maximization algorithm to crawl along the multi-dimensional functional surface while searching for the point where the signature recreated from library elements most closely approximates the unknown signal. To attempt to streamline the process, we take the results of the correlation coefficient algorithm and transfer them to the linear combination program thus giving the program an a priori knowledge of the components it should attempt to maximize. The results of this algorithm are summarized in Figure 3 where we compare the fractional molecular abundance expected by knowledge of the mixture components to that reconstructed by our algorithm.

Linear Combination Limitations

Examination of the the method utilized for deconvolution of the mixtures has revealed some limitations. Namely and undercounting of the water molecules present, a propensity for confusion between similar library samples and an unexplained behavior in the fractional abundance of acetonitrile which bears further examination. Analysis of the results of Figure 3 show that in each instance the fractional abundance of water was undercounted when compared to that which should appear in the mixture. We attribute this to the reduced absorbance of water as shown in Figure 4 when compared to the other chemicals. The reduced absorbance of water should result in a larger collection volume during the examination of the library signal and thus a larger signal than is present within the collection volume present in the mixture sample.

Further examination of the results shows that the correlation algorithm missed a component in the 5 chemical mixture. Through both the correlation study as well as the creation of the demixing algorithm we noticed that there are many similarities between the Raman spectra of Ethanol and Ethylene Glycol, unsurprising given the similarities in certain bond structures. This non-orthogonality within the library set resulted in the algorithms led to very small distinguishing differences between the weighting scores of several mixture variations. In this instance, the abundance of Ethylene Glycol was approximately twice what would have been expected and subsumed the contribution from the Ethanol within the mixture.

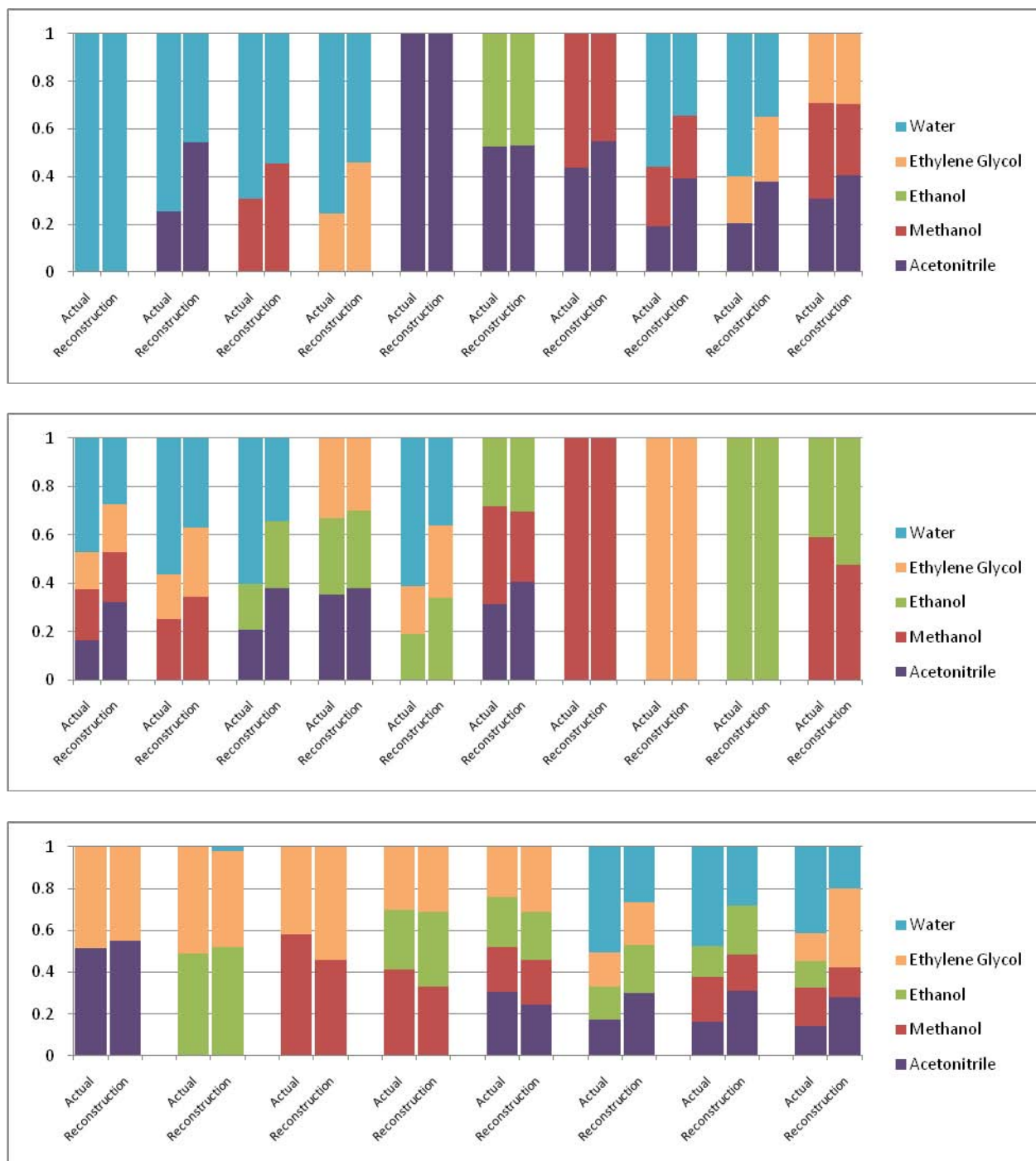


Figure 3: Fractional Chemical Abundances, in each pair of bars the bar on the left represents a calculation of the actual fractional chemical abundance based upon molecular mass and density and the column on the right is the reconstructed value obtained from the demixing algorithm.

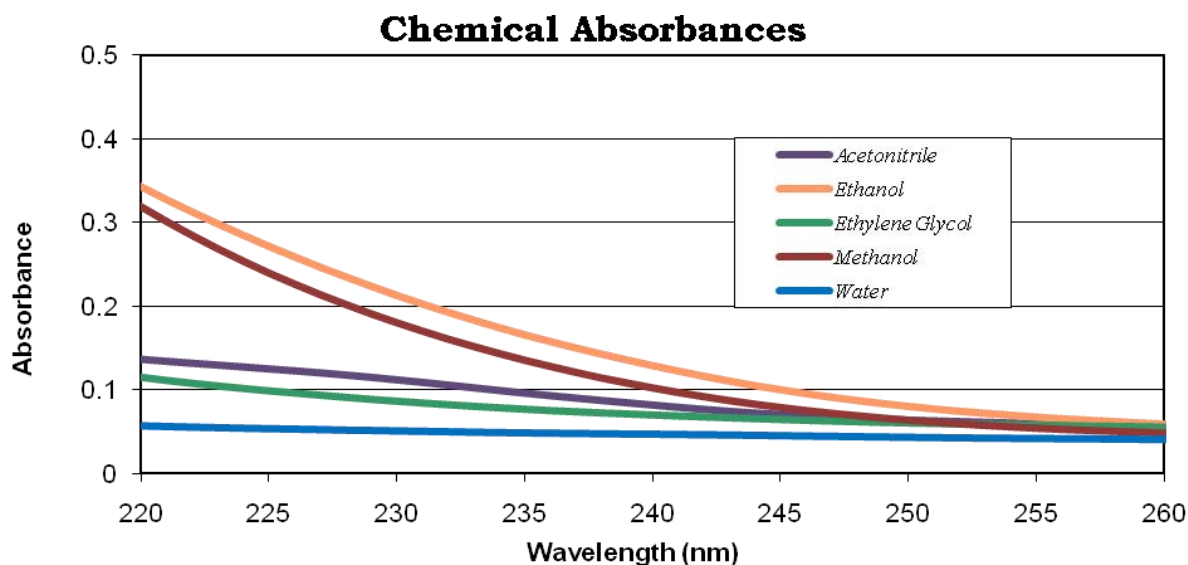


Figure 4: Absorbance of the library chemicals over the selected wavelength range.

Conclusion

Utilizing a fully contained mixture set we have measured both the components and the possible mixture combinations of those components in the ultraviolet range from 220 – 260 nm utilizing the SWOrRD spectroscopy system. We have demonstrated that reliable identification of mixture components is possible and that a reasonable reconstruction of the fractional abundance of the constituent chemicals is possible while utilizing a linear mixture model. We will further look into the effects of the absorbance of a mixture in the undercounting of components and also the variability of mixture weights within the wavelength range.

Acknowledgement

This work is sponsored by The Defense Threat Reduction Agency (DTRA). DTRA was founded in 1998 to integrate and focus the capabilities of the Department of Defense that address the weapons of Mass Destruction (WMD) threat. The mission of the DTRA is to safeguard America and its allies from WMD

threat (e.g. chemical, biological, radiological, nuclear and high yield explosives) by providing capabilities to reduce, eliminate and counter the threat and mitigate its effects. Under DTRA, Department of Defense resources, expertise and capabilities are combined to ensure the United States remains ready and able to address the present and future WMD threats. For more information on DTRA, visit www.dtra.mil/.